

The Rise and Fall of R&D Networks

Mario V. Tomasello,^{*} Mauro Napoletano,[†]
Antonios Garas[‡] and Frank Schweitzer[§]

April 15, 2013

Abstract

Drawing on a large database of publicly announced R&D alliances, we track the evolution of R&D networks in a large number of economic sectors over a long time period (1986–2009). Our main goal is to evaluate temporal and sectoral robustness of the main statistical properties of empirical R&D networks. By studying a large set of indicators, we provide a more complete description of these networks with respect to the existing literature. We find that most network properties are invariant across pooled and sectoral networks. This result suggests the presence of universal laws governing the dynamics of R&D networks. Moreover, we find that many properties of R&D networks are characterized by a non-monotonic trend with a peak in the mid-nineties. This allows us to define the mid-nineties as a “Golden Age” characterized by the rise-and-fall of R&D networks. Finally, we show that many properties of empirical R&D networks support predictions of the recent theoretical literature on R&D network formation.

1 Introduction

Strategic alliances and joint Research and Development (R&D) activities between firms are recognized as an important factor to foster technology innovation and economic growth. In particular, the 1990s witnessed an unprecedented growth of such activities (e.g. Ahuja, 2000; Gulati, 1995; Hagedoorn, 2002). In this study, R&D alliances include any kind of agreement between two firms that is aimed at R&D purposes and can be traced in public announcements or press releases. Our work is related to the growing body of empirical literature that has studied R&D alliances using the approach of *complex networks*. In this approach, nodes represent firms and links represent their R&D alliances. The network evolves over time both with respect to the number of nodes and the number of links (see Bojanowski et al., 2012; Powell et al., 1996; Verspagen and Duysters, 2004).

^{*}Corresponding author. E-mail: mtomasello@ethz.ch. Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland.

[†]OFCE and SKEMA Business School, Sophia-Antipolis, France, and Sant’Anna School of Advanced Studies, Pisa, Italy. Mail Address: OFCE, (Bâtiment 2), 250 rue Albert Einstein, 06560 Valbonne, France. E-mail: mauro.napoletano@ofce.sciences-po.fr.

[‡]Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland. E-mail: agaras@ethz.ch.

[§]Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland. E-mail: fschweitzer@ethz.ch.

Several works have tried to shed light on the structural properties of R&D networks. It was shown that they are typically sparse and characterized by heavily asymmetric degree distributions (e.g. Hanaki et al., 2007, 2010; Powell et al., 2005b; Rosenkopf and Schilling, 2007). As another prominent feature, R&D networks can be described as small worlds (e.g. Fleming et al., 2007; Fleming and Marx, 2006), i.e. networks displaying short average path length and high clustering (Watts and Strogatz, 1998). However, it has been recognized that the observed R&D network structures oscillate over time. That means, these networks exhibit a periodical evolution of properties such as size, density and centralization (i.e. the concentration of node degree or other centrality measures). One prominent example is the rise and fall of small world properties that Gulati et al. (2012) has found in the R&D network of the global computer industry.

The above literature has greatly contributed to the understanding of empirically observed R&D networks. At the same time, empirical studies have often focused only on a small number of industries or have rarely considered the properties of the networks at different time periods. Finally, they have considered only a limited set of network measures (e.g. size, degree heterogeneity, small world property). We improve upon the foregoing literature along several dimensions. *First*, we analyze the R&D networks in a large number of manufacturing, service and non-industrial (e.g. universities) sectors. We study the pooled R&D network, i.e. the network containing all alliances independently of the sectors to which the partners belong, comparing it to a disaggregated analysis of the network into industrial sectors at a 3-digit SIC level. Via this disaggregated analysis, we are able to check whether the network properties that have been analyzed by the current literature for sectors like computers (e.g. Hanaki et al., 2010) or pharmaceuticals (e.g. Powell et al., 2005a) are robust across different sectors of activity. In addition, by comparing the properties at the pooled and at the sectoral levels, we are able to check for the presence of statistical regularities in R&D networks that hold irrespectively of the scale of aggregation at which they are observed. *Second*, we perform a longitudinal analysis of empirical R&D networks. In particular, we consider the network evolution in the period from 1986 to 2009, and we study network properties in 6 different sub-periods (1986-1989, 1990-1993, 1994-1997, 1998-2001, 2002-2005, 2006-2009). This procedure allows us to check whether network properties are robust over time, or if instead they exhibit different patterns in each considered time-period. *Third*, we investigate a broad set of network properties. We start our analysis by studying basic measures of the network (size, degree heterogeneity), as well as measures related to small world properties (path length and clustering). In addition, we study measures related to more complex features of the network, such as assortativity or the emergence of community and hierarchical structures. This way, we extend the existing knowledge of R&D networks by adding new stylized facts; furthermore, we provide confirmation to some of the predictions of the theoretical literature on R&D networks. Indeed, recent theoretical works have shown that – under non-negligible costs of collaboration – “efficient” networks (i.e. networks maximizing some measure of industry welfare) are typically asymmetric, and organized into hierarchical core-periphery structures (see e.g. Goyal and Joshi, 2003; König et al., 2012; Westbrock, 2010). These works have also shown that the networks emerging under endogenous partner selection can be too little asymmetric or too

little connected with respect to what is socially desirable¹ (see König et al., 2012; Westbrock, 2010).

We find that both the pooled and the sectoral networks exhibit the same patterns for a wide set of network properties, like the fraction of firms in the largest connected component, the shape of degree distributions or the presence of small world properties. In addition, both the pooled and the sectoral R&D networks are organized into core-periphery architectures characterized by a high degree of nestedness. In contrast, pooled and sectoral networks differ with respect to assortativity. In the pooled R&D network firms with many alliances tend to collaborate with partners involved in many alliances as well, whilst the opposite can be found at the sectoral level, where firms with many alliances tend to collaborate with firms having fewer alliances. Furthermore, all of the above properties display a non-monotonic trend over time. For instance, the network size (related to the number of R&D alliances) of the pooled network has first increased over time, reaching a peak in the period 1994-1997, and then it has significantly decreased until the end of our observation period. As we show in Section 3, this dynamics was driven by a growth in the number of firms participating in R&D alliances rather than by the change in the number of alliances among firms already involved in previous collaborations. Interestingly, the above dynamics is very pervasive, as we observe it in most of the sectoral R&D networks we study.

The above results have several implications for the literature. First, the finding that both the pooled and the sectoral R&D networks are characterized by core-periphery and nested architectures is in line with some predictions of the recent theoretical literature on R&D networks. Second, the existence of features that are invariant with respect to the scale of aggregation favors the idea that the laws governing R&D network formation can be analyzed independently of the characteristics of the sector to which firms belong. Finally, our results show that the non-monotonic trend, which has so far been emphasized only in relation to network size and small worlds, is also displayed by more sophisticated topological network properties such as core-periphery and nestedness, that are not trivially related to the network size, but to more fundamental underlying mechanisms. Extending the definition proposed by Gulati et al. (2012), we argue that such mechanisms lead to the *rise-and-fall* of R&D networks, both at the pooled and at the sectoral levels. We show, by means of quantitative indicators, how R&D networks arise, go through their “golden age” and finally decline.

The paper is organized as follows. In section 2 we describe the database and the methodology employed to analyze the pooled and the sectoral R&D networks. In section 3 we discuss results about the basic properties of R&D networks, such as network size and the fraction of firms in the largest connected component. In section 4 we analyze the characteristics of the degree distributions. In sections 5 and 6 we study more sophisticated network properties, such as assortativity, small worlds and communities in the network and, finally, the emergence of core-periphery architectures. In section 7 we draw our conclusions.

¹ König and Tessone (2009) show that core-periphery structures can also be possible equilibrium structures when firms try to form alliances with the most central firms in the networks

2 Data and Methodology

An *R&D network* is a representation of the interactions occurring between firms in one or more industrial sectors and in a given period of time. Every network consists of a set of *nodes* and *links* connecting pairs of nodes. In our representation, each node of the network is a *firm* and every link represents an *R&D alliance* between two firms. By R&D alliance, we refer to an event of partnership between two firms, that can span from formal joint ventures to more informal research agreements, specifically aimed at research and development purposes. To detect such events, we use the *SDC Platinum* database, provided by Thomson Reuters, that reports all publicly announced alliances, from 1984 to 2009, between several kinds of economic actors (including manufacturing firms, investors, banks and universities). We then select all the alliances characterised by the “R&D” flag; after applying this filter, a total of 21572 alliances are listed in the dataset. Information in the SDC dataset is gathered only from announcements in public sources, such as press releases or journal articles. Nevertheless, despite the bias that could be introduced by such a collection procedure, the work by Schilling (2009) shows that the SDC Thomson dataset provides a consistent picture with alternative alliance databases (e.g. CORE and MERIT-CATI) in terms of alliance activity over time, geographical location of companies and industry composition.

Subsequently, we draw a link connecting two nodes every time an alliance between the two corresponding firms is announced in the dataset. An alliance is associated with an *undirected* link, as we do not have any information about the initiator of the alliance. When an alliance involves more than two firms (*consortium*), all the involved firms are connected in pairs, resulting into a fully connected clique. Multiple links between the same nodes are in principle allowed (two firms can have more than one alliance on different projects). Nevertheless, as we aim at studying the connections between firms, and not the number of alliances a firm is involved in, we discard this information and use *unweighted* links in our network representation. For this reason, we define the *degree* of a node as the number of other nodes to which it is linked, i.e. the number of partners that a firm has – not the number of alliances. Furthermore, a firm appears in the R&D network only if it is involved in at least one alliance. Our study is focused exclusively on the embeddedness of firms in an alliance network. For this reason, isolated nodes are not part of our network representation.

Both the links and the nodes of the R&D network are characterized by a *entry/exit dynamics*. Alliances between firms are shown to have a finite duration (see Deeds and Hill, 1999; Phelps, 2003). This causes some firms to disappear from the network, after they no longer participate in any alliance. Likewise, many new firms that were not listed in any previous alliance may enter the network at the beginning of a new year. Our longitudinal study clearly requires precise temporal information about the formation and the deletion of alliances. The SDC Platinum dataset contains the beginning date of every alliance, but there is no information about any of the ending dates (firms do usually not organize press releases to announce the cut of an

alliance). We are thus forced to make some assumptions about the alliance durations. We started by drawing the duration of every alliance from a normal distribution with mean value from 1 to 5 years and standard deviation from 1 to 5 years. We observed that all our results remain qualitatively unchanged by changing the mean value and the standard deviation within these ranges. The variation of the standard deviation has nearly no influence on the trend of all the measures we compute on the networks. The variation of the mean alliance duration changes the absolute values of the network indicators while not affecting their trend and peak positions.² Given the strong robustness of the R&D network to the variation of alliance lengths, we took a conservative approach and assumed a fixed 3-year length for every partnership, consistently with previous empirical work (e.g. Deeds and Hill, 1999; Phelps, 2003).

As the SDC Platinum dataset does not have a unique identifier for each firm, all the associations between alliances and firms (i.e. the construction of the network itself) are based only on the firm names reported in the dataset. Thus, it could be the case that two or more entries are listed with different names, because they appear in two distinct alliance events, even though they correspond to the same firm. For this reason, we checked all firm names and control for all legal extensions (e.g. 'ltd', 'inc', etc.) and other recurrent key words (e.g. 'bio', 'tech', 'pharma', 'lab', etc.) that could affect the matching between entries referring to the same firm. We decided to keep as separated entities the subsidiaries of the same firm located in different countries. The raw dataset contains a total of 16313 firms, which are reduced to 14561 after running such an extensive standardization procedure.

The rules we used to construct the R&D network are the following: we link two nodes when an alliance between the corresponding firms occurs and we delete this link 3 years after its formation. In this way, we are able to build 26 snapshots of the R&D network – one for every year – from 1984 to 2009. From now on we call this network, containing all companies irrespective of their industrial sector, the *pooled R&D network*.

Every firm listed in the SDC Platinum dataset is associated with its SIC (Standard Industrial Classification), a US-government code system for classifying industrial sectors. This allowed us to build the *classified R&D networks* for the several sectors that we can identify in the dataset. A classified R&D network centered around a given sector contains only alliances for which at least *one* of the two firms has a three-digit SIC code matching the selected sector.³ The rules for link deletion are the same as in the pooled R&D network. More precisely, we selected for our study the 30 largest industrial sectors, in terms of number of firms engaged in alliances in 1995 (the year in which the pooled R&D network reached its maximum size). This list includes man-

²When the mean alliance duration is further increased we observe the emergence of inertia and accumulation effects. To give an extreme example, when the mean alliance duration is 15 years, the network size shows a monotonic growing trend, because alliances between firms are hardly ever cut. Nevertheless, even in this unrealistic scenario, relative measures such as average degree or giant component fraction exhibit not only qualitatively equal trends, but also unchanged peak positions, meaning that our analysis is resilient to alliance duration changes and gives consistent results on firms' alliance activity.

³This approach is used also by Rosenkopf and Schilling (2007).

ufacturing sectors (Pharmaceuticals, Electronic Components, Medical Supplies, Aircrafts and so on), but also services (Computer Software, Investors) and knowledge-based sectors (Laboratories and Testing companies, Universities).

We studied both the pooled R&D network and the classified R&D networks by computing a set of network indicators along the whole observation period, with the exception of the first two years, when the networks are too small – or not even existing for some industries. All results are presented below. We group our analysis into four sections: basic network statistics, heterogeneity in alliance behavior, small world and communities, core-periphery and nested architectures.

3 Basic Network Statistics

In Fig. 1 six snapshots of the pooled R&D network, in 1989, 1993, 1997, 2001 and 2005 are shown. The plots are produced using the R package *igraph*, and the networks are displayed using the Fruchterman-Reingold algorithm. This is a force-based algorithm for network visualisation which positions the nodes of a graph in a two-dimensional space so that all the edges are of similar length and there are as few crossing edges as possible. The result is that the most interconnected nodes are displayed close to each other in the two-dimensional plot. The ten largest industrial sectors are depicted with different colors. The figure shows that two clusters always dominate the pooled R&D network: a cluster centered on pharmaceutical companies and a cluster centered on *ICT*-related companies.

Fig. 1 already suggests the existence of a cyclic fluctuation in the size of the pooled network of R&D alliances. To shed more light on this, we report in Table 1 the network size, in terms of number of firms taking part in the R&D network – i.e. companies involved in at least one alliance. The observation time period 1986-2009 has been divided into six time sub-periods of 4 years each and we averaged the network size within each of the sub-periods. The number of companies involved in R&D alliances increases to a peak in the mid-nineties and then shrinks again, both at the pooled and the sectoral level. Such a non-monotonic trend holds for all the sectoral R&D networks: in each of them, the number of firms involved in R&D alliances has a peak in the years 1994-1997. Interestingly, only the Pharmaceutical sector besides the peak in the period 1994-1997, has an additional peak of slightly larger size in the period 2006-2009.

Next, we computed the fraction of nodes belonging to the largest *connected component* of the network. A connected component is defined as a set of nodes which are connected to each other by at least one path (i.e. a sequence of links). We refer to the largest connected component as the *giant component* of the network. The giant component size to the overall network size ratio (or *giant component fraction*) is a rough indicator of the network connectedness. Our results are reported in Table 2. This measure has been computed for every year from 1986 to 2009 and then averaged within six time sub-periods of 4 years each. Similar to the network size, the giant component fraction displays a non-monotonic trend at the pooled level, reaching a peak in the

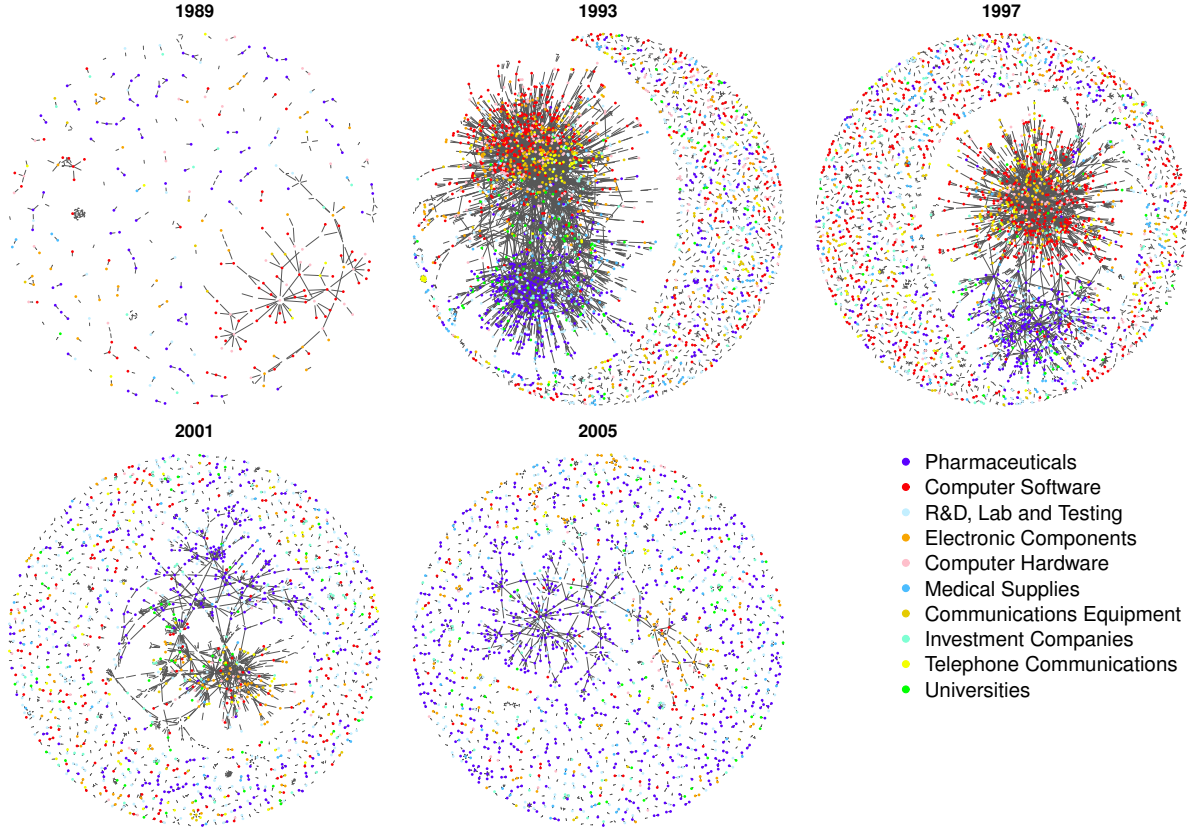


Figure 1: Pooled R&D network snapshots in 1989, 1993, 1997, 2001 and 2005. We plotted – in different colours – only the twelve largest sectors, in order to ease visualisation.

mid-nineties and then shrinking again. Fig. 1 gives a visual idea of this evolution: the pooled R&D network is dominated by a giant component that expands until 1997 and subsequently leaves space to a growing periphery of disconnected dyads (pairs of partnered firms). The emergence of a giant component in the network is of particular interest, as different theoretical works (e.g. Goyal and Joshi, 2003; König et al., 2012) have stressed the importance of the relation between high network connectedness and efficiency in terms of aggregate profits. Moreover, we find that the emergence of such non-monotonic trend in the giant component is very robust to sectoral disaggregation. Indeed, we observe it in almost all the sub-networks representing the different industrial sectors (see Table 2). It has to be noticed that the early observation period 1986-1989 is characterised by very high giant component fractions, because of spurious effects due to extremely small network size. After discarding these observations, the pooled R&D network exhibits a giant component peak from 1990 to 1997; moreover, 19 out of the 30 sectoral R&D networks show a giant component peak either in the 1990-1993 or in the 1994-1997 period. The sectors that do not have a peak show a more volatile evolution of their giant component; among these, only 4 are manufacturing industries (Inorganic Chemicals, Household Audio-Video, Special

Table 1: Network size for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). The values are averaged within 6 time sub-periods of 4 years each.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	280	2515	4918	2626	2219	1829
Computer Software (737)	69	560	1488	549	284	122
Pharmaceuticals (283)	77	645	935	682	825	949
R&D, Lab and Testing (873)	26	477	848	534	596	500
Computer Hardware (357)	51	385	744	202	92	29
Electronic Components (367)	54	328	581	253	222	165
Communications Equipment (366)	17	207	475	181	113	60
Medical Supplies (384)	10	164	280	122	119	123
Telephone Communications (481)	12	184	350	132	82	22
Universities (822)	3	192	374	166	152	83
Laboratory Apparatus (382)	10	139	243	116	94	87
Investment Companies (679)	14	138	298	232	207	125
Motor Vehicles (371)	6	108	190	97	85	78
Aircrafts and parts (372)	8	83	136	60	40	26
Professional Equipment Wholesale (504)	4	64	142	26	8	8
Inorganic Chemicals (281)	15	108	152	50	45	31
Engineer.,Architec.,Survey (871)	2	74	129	62	26	16
Household Audio-Video (365)	9	110	164	90	65	30
Plastics (282)	11	97	121	44	36	18
Motion Picture Production (781)	-	15	91	14	4	1
Management,Consulting,PR (874)	1	28	96	61	64	28
Electrical Machinery NEC (369)	2	54	96	26	24	37
Radio and TV Broadcasting (483)	2	26	88	22	7	4
Cable and TV Services (484)	-	18	78	8	6	3
Special Machinery (355)	2	33	82	34	17	11
Business Services (738)	1	15	66	37	30	5
Crude Oil and Gas (131)	3	42	72	62	35	27
Electrical Goods Wholesale (506)	-	26	84	19	10	8
Naut./Aeronaut. Navigation (381)	1	49	82	21	16	12
Electric Services (491)	-	50	78	38	26	15
Organic Chemicals (286)	5	44	60	18	23	18

Machinery, Organic Chemicals), while the other sectors are related to services or sales.

The analysis of the first network measures – size and giant component fraction – reveals the existence of trends that are invariant to the scale of aggregation or the sector where they are observed. In particular, the years between 1990 and 1997 can be thought as a “golden age” of R&D networks, witnessing not only a higher number of alliances, but also the emergence of a giant component. In the next section, we will go into more detail on how these alliances are organized, by studying the characteristics of the degree distributions of the pooled and sectoral R&D networks.

Table 2: Fraction of the giant component for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). The values are averaged within 6 time sub-periods of 4 years each. High values in the first sub-period 1986-1989 are affected by very small network sizes.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	0.10	0.53	0.53	0.33	0.26	0.20
Computer Software (737)	0.33	0.54	0.54	0.23	0.11	0.06
Pharmaceuticals (283)	0.08	0.58	0.68	0.49	0.36	0.32
R&D, Lab and Testing (873)	0.13	0.19	0.27	0.11	0.10	0.07
Computer Hardware (357)	0.27	0.59	0.67	0.51	0.28	0.13
Electronic Components (367)	0.15	0.53	0.61	0.49	0.38	0.13
Communications Equipment (366)	0.18	0.42	0.55	0.25	0.25	0.15
Medical Supplies (384)	0.21	0.04	0.05	0.05	0.06	0.05
Telephone Communications (481)	0.43	0.61	0.58	0.25	0.26	0.28
Universities (822)	0.90	0.17	0.25	0.10	0.08	0.05
Laboratory Apparatus (382)	0.26	0.15	0.13	0.08	0.08	0.07
Investment Companies (679)	0.21	0.36	0.27	0.23	0.28	0.10
Motor Vehicles (371)	0.79	0.52	0.39	0.15	0.21	0.10
Aircrafts and parts (372)	0.65	0.47	0.38	0.23	0.20	0.16
Professional Equipment Wholesale (504)	0.69	0.13	0.16	0.23	0.37	0.28
Inorganic Chemicals (281)	0.30	0.26	0.17	0.15	0.12	0.29
Engineer.,Architec.,Survey (871)	1.00	0.12	0.15	0.11	0.12	0.20
Household Audio-Video (365)	0.61	0.57	0.61	0.63	0.60	0.28
Plastics (282)	0.23	0.25	0.20	0.23	0.15	0.19
Motion Picture Production (781)	-	0.39	0.24	0.22	0.62	0.50
Management,Consulting,PR (874)	1.00	0.23	0.07	0.09	0.09	0.11
Electrical Machinery NEC (369)	1.00	0.36	0.22	0.20	0.15	0.11
Radio and TV Broadcasting (483)	1.00	0.40	0.17	0.16	0.42	0.61
Cable and TV Services (484)	-	0.35	0.16	0.31	0.53	0.75
Special Machinery (355)	0.88	0.25	0.13	0.19	0.27	0.26
Business Services (738)	1.00	0.48	0.08	0.11	0.14	0.65
Crude Oil and Gas (131)	0.67	0.15	0.14	0.10	0.11	0.15
Electrical Goods Wholesale (506)	-	0.29	0.12	0.15	0.25	0.34
Naut./Aeronaut. Navigation (381)	1.00	0.38	0.26	0.21	0.22	0.24
Electric Services (491)	-	0.35	0.11	0.15	0.24	0.21
Organic Chemicals (286)	0.73	0.13	0.17	0.25	0.13	0.22

4 Heterogeneity in alliance behavior

4.1 Degree Distributions

A large part of literature has analyzed some properties of the degree distributions in R&D networks. Empirical studies have shown that degree distributions in R&D networks tend to be highly skewed. Some studies report exponential distributions (Riccaboni and Pammolli, 2002), while others find power-law distributions (Powell et al., 2005b). The presence of a power-law distribution would indicate the existence of an underlying multiplicative growth process (Reed, 2001; Simon, 1955). In the context of R&D networks this means that firms which have many collaborations already attract more new partners than firms with only few collaborations. This idea underlies the “preferential attachment” model by Barabasi and Albert (1999). However, this model assumes that firms know how many collaborations every other firm in the network has. This may become unrealistic, especially in large networks or situations in which this information is

not publicly available. More realistic models assume that firms have only local information about the network, for example about their current alliance partners and their collaborations, that is, their neighbors' neighbors. The network formation model introduced by König et al. (2012) and König and Tessone (2009) assumes that firms search for the most central partner in their local neighborhood. Their model generates exponential degree distributions with power-law tails. Another model in which agents form links locally is the one by Jackson and Rogers (2007), that generates power-law degree distributions as well as exponential degree distributions, depending on various parameters. We extend the existing discussion about the degree distributions in R&D networks by studying their evolution over time and comparing the results between different sectors. Given the small size of many of our networks, we did not test or validate any functional form, but we rather measured the statistical properties of the degree distributions, in order to assess their main features and get insights into the underlying network formation process. We state the following hypothesis:

Hypothesis 4.1. *R&D networks across sectors and time are characterized by dispersed and skewed degree distributions. Periods with higher alliance activity exhibit more dispersed, skewed and fat-tailed degree distributions.*

As in the previous section, the whole observation period is divided into six sub-periods lasting 4 years, and all the observations of firms' degree are pooled within each sub-period, for both the pooled R&D network and the sectoral R&D networks. All the measures we present in this section were computed for these pooled networks, in each of the time sub-periods. The degree distribution of the pooled R&D network in the six time sub-periods is shown in Fig. 2. As already mentioned in Sec. 2, we define the degree as the number of partners of a firm, and not the number of alliances. For this reason, we counted multiple alliances between the same two firms as one, and we counted all the firms participating in the same consortia as distinct partners. Let us define $p(x)$, commonly called *probability density function*, as the fraction of nodes in the network with degree x . What we plot in Fig. 2 is not the probability density function of our degree distributions, but the *complementary cumulative distribution function* $P(x)$, defined as the fraction of nodes having degree greater than or equal to x :

$$P(x) = \int_x^{\infty} p(x') dx'. \quad (1)$$

The visual form of the complementary cumulative distribution function is more robust than that of the probability density function against fluctuations due to finite sample sizes, particularly in the tail of the distributions. We observe that the degree distribution of the pooled R&D network is very broad and skewed, during all the observation period. Moreover, the shape of the degree distribution is independent of the network size. For instance, the degree distributions of the pooled R&D network in the “golden age” 1994-1997 (maximum degree ~ 200) has a very similar trend to that of the early period 1986-1989 (maximum degree ~ 20). In addition, most of

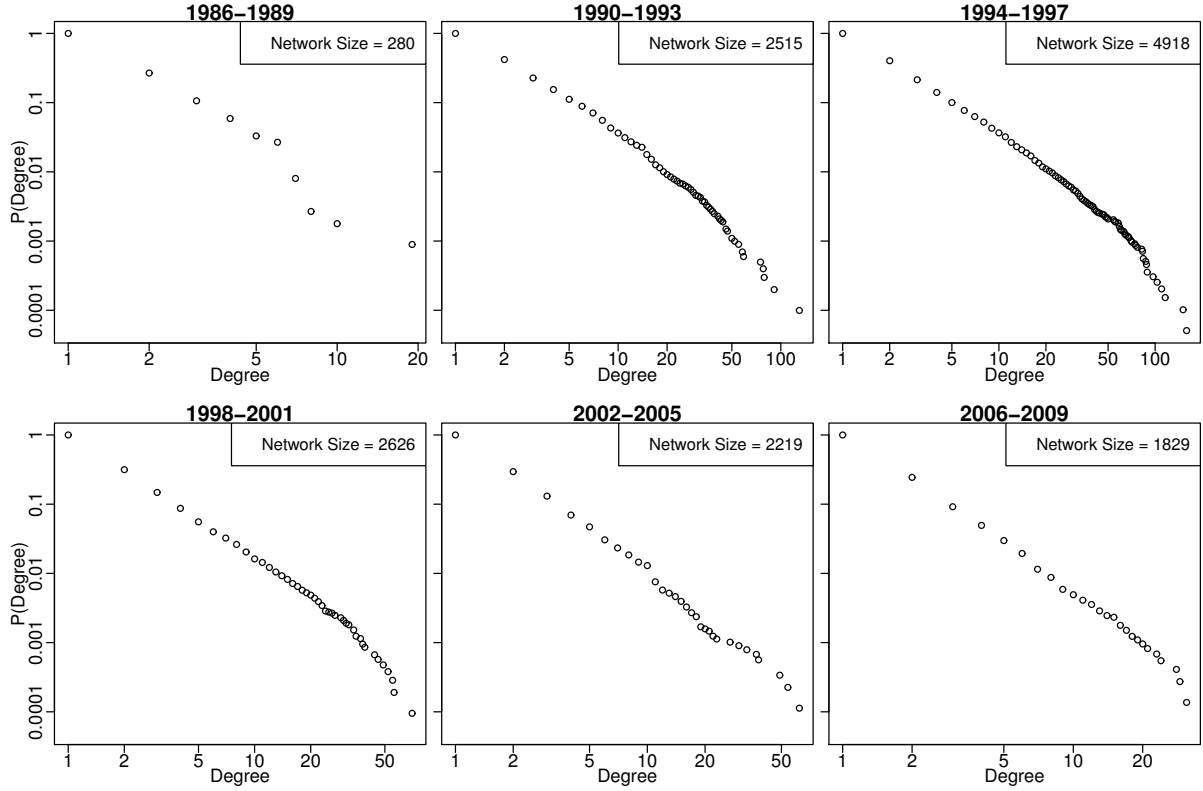


Figure 2: Complementary cumulative degree distribution for the pooled R&D network. Observations are aggregated in six time sub-periods.

the sectoral R&D networks exhibit this kind of degree distribution, during the whole observation period.⁴

We present in Table 3 a summary of the statistics for the pooled R&D network degree distribution, showing the first four moments (mean value, standard deviation, skewness and kurtosis), in the same time periods as those shown in Fig. 2. The normality Kolmogorov-Smirnov test shows that the degree distributions are extremely far from the Normal benchmark, during the whole observation period: all of them are indeed characterized by a great variance and skewness, with the presence of a fat tail – a small number of firms having a huge number of partners. The distributions seem to be power-laws, but – as already said – we do not aim at validating any functional form; we limit ourselves to report the most relevant statistical properties. Each of the first four moments of the degree distribution increases to reach a peak either in 1990-1993 or in 1994-1997 and then decreases again. Even though the arithmetic mean is not entirely meaningful or predictive for power-law-like distributions, we still report here this measure not only because it is

⁴For the sake of simplicity, we do not plot here the degree distributions for all the sectoral R&D networks. More detailed data and plots are available upon request.

fully computable (we have finite size networks), but also because it gives an idea about the firms' alliance activity during our observation period. The mean degree has a value of 1.51 partners per firm in the early period 1986-1989; it then exhibits a peak value in 1990-1993 (2.52 partners per firm), which remains almost unchanged in 1994-1997 (2.51 partners per firm), showing that firms have on average more alliance partners in the "golden age" of alliance formation. The average number of partners per firm eventually decreases again, reaching a value of 1.49 in the late period 2006-2009.

The degree distribution in the pooled R&D network is highly dispersed, as shown by standard deviation values that are always comparable or even larger than the mean values. This holds especially for the 1994-1997 period, when the standard deviation has a peak at 4.98, while the mean value is 2.51 partners per company. The skewness is always positive, in particular during the "golden age" (the peak is reached once again in 1994-1997); this suggests the presence of a heavy right tail, consisting of a small amount of firms having a huge number of partners. Finally, all the degree distributions show *leptokurtosis*, which means that their kurtosis is much bigger than the expected value of a normal distribution (equal to 3). The kurtosis reaches its peak in 1994-1997; this is again indicative of heavy tails in the R&D networks degree distributions.

Table 3: Degree distribution statistics and p -Values of Kolmogorov-Smirnov (KS) test for the pooled R&D network. Observations are aggregated in six time sub-periods.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Mean	1.51	2.52	2.51	1.87	1.70	1.49
SD	1.22	4.30	4.98	2.77	2.11	1.45
Skewness	4.90	9.35	11.28	9.26	10.56	7.92
Kurtosis	47.30	158.40	206.69	133.70	200.25	104.84
KS test p -Value	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$

The degree distributions of the sectoral R&D networks display patterns that are similar to those of the pooled R&D network. We report in Table 4 the values of the average degree for the pooled and the sectoral R&D networks in 6 time periods, that clearly confirm such a cross-sector similarity. In all the sectoral networks, firms have on average more collaborators during the "golden age" of alliance activity (1990-1993 and 1994-1997 periods). The only two exceptions are represented by two manufacturing industries, motor vehicles (having a peak in 1986-1989) and organic chemicals (that has a first peak in 1986-1989 and a second one in 1994-1997).

As another cross-sector similarity, we now study the skewness of the degree distributions in the sectoral R&D networks. As already mentioned, this quantity is always positive during the whole observation period, confirming the strong asymmetry of the degree distributions. In order to get an estimate about the tail of the degree distributions from a non-parametric point of view, we compute the Hill Estimator (Hill, 1975), a tool commonly used to study the tails of economic data. If n is the number of observations (in our case, the number of nodes in the R&D network) and k is the number of tail observations ($k \leq n$), the inverse of the Hill estimator (HE) is defined

Table 4: Average degree (number of partners) for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). Observations are aggregated in six time sub-periods. Missing values refer to empty networks in the early period 1986-1989.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	1.51	2.52	2.51	1.87	1.70	1.49
Computer Software (737)	1.70	2.16	2.21	1.52	1.27	1.13
Pharmaceuticals (283)	1.22	2.09	2.22	1.82	1.57	1.55
R&D, Lab and Testing (873)	1.08	1.68	1.81	1.40	1.43	1.27
Computer Hardware (357)	1.50	2.10	2.45	2.30	1.55	1.09
Electronic Components (367)	1.32	2.18	2.38	2.15	1.81	1.44
Communications Equipment (366)	1.10	1.82	2.03	1.57	1.48	1.34
Medical Supplies (384)	1.00	1.26	1.31	1.21	1.20	1.16
Telephone Communications (481)	1.19	2.84	2.53	1.42	1.57	1.28
Universities (822)	1.27	1.66	1.76	1.51	1.35	1.11
Laboratory Apparatus (382)	1.00	1.41	1.36	1.24	1.20	1.19
Investment Companies (679)	1.04	1.74	1.62	1.53	1.63	1.35
Motor Vehicles (371)	2.31	1.89	1.78	1.40	1.49	1.29
Aircrafts and parts (372)	2.00	2.25	2.00	1.68	1.41	1.40
Professional Equipment Wholesale (504)	1.22	1.24	1.42	1.22	1.09	1.00
Inorganic Chemicals (281)	1.28	1.48	1.53	1.23	1.17	1.27
Engineer.,Architec.,Survey (871)	1.00	1.36	1.40	1.17	1.07	1.09
Household Audio-Video (365)	1.44	2.11	2.61	2.32	2.20	1.58
Plastics (282)	1.07	1.54	1.55	1.46	1.29	1.11
Motion Picture Production (781)	-	1.38	1.36	1.02	1.00	1.00
Management,Consulting,PR (874)	1.00	1.20	1.20	1.19	1.16	1.06
Electrical Machinery NEC (369)	1.00	1.45	1.52	1.26	1.11	1.10
Radio and TV Broadcasting (483)	1.33	1.69	1.31	1.15	1.11	1.11
Cable and TV Services (484)	-	1.34	1.51	1.03	1.17	1.00
Special Machinery (355)	1.00	1.34	1.37	1.24	1.21	1.07
Business Services (738)	1.00	1.17	1.22	1.15	1.16	1.05
Crude Oil and Gas (131)	1.09	1.70	1.68	1.51	1.28	1.11
Electrical Goods Wholesale (506)	-	1.35	1.34	1.06	1.05	1.07
Naut./Aeronaut. Navigation (381)	1.33	1.49	1.49	1.23	1.13	1.09
Electric Services (491)	-	1.57	1.38	1.22	1.22	1.25
Organic Chemicals (286)	1.26	1.17	1.26	1.14	1.09	1.12

as:

$$\hat{h}^{-1} = k^{-1} \sum_{i=1}^k [\log(x_i) - \log(x_{min})], \quad (2)$$

where x_i , $i = 1 \dots k$ are the tail observations, i.e. the degree values such that $x_i \geq x_{min}$. The smaller the HE value, the “heavier” the tail of the degree distribution is. As reported in Table 5, the HE is smaller than 6.3 for all sectors in all time periods, a value that confirms the presence of heavy tails in the degree distributions. In particular, the HE is smaller than 4 for all the sectors in the “golden age” (1994-1997), suggesting that such tails in the degree distributions are even heavier when the alliance activity is higher.⁵

⁵The Hill estimator is defined as a non-parametric estimate for the “heaviness” of the tails of a distribution. Nevertheless, if the data are power-law-like (or Pareto-like) distributed, as evidence suggests in our case (see Fig. 2), the interpretation of the HE might be easier from a parametric perspective. The HE is indeed shown to be equivalent to the maximum-likelihood estimator of the parameter α in a Pareto distribution, whose cumulated density is $P(x) = 1 - x^{-\alpha}$.

Table 5: Hill estimator (HE) for degree distributions in the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). Observations are aggregated in six time sub-periods. Missing values of HE refer to networks with too few observations; all the sectoral R&D sub-networks have a significant number of observations in the “golden age”.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	3.04	2.31	2.34	2.61	2.78	3.05
Computer Software (737)	2.71	2.41	2.30	2.70	3.31	4.24
Pharmaceuticals (283)	5.19	2.91	2.45	2.58	2.89	3.02
R&D, Lab and Testing (873)	-	2.77	2.69	3.65	3.23	3.60
Computer Hardware (357)	2.70	2.37	2.22	2.75	2.88	4.59
Electronic Components (367)	3.36	2.43	2.43	2.25	2.59	3.57
Communications Equipment (366)	-	2.66	2.50	2.43	2.71	2.65
Medical Supplies (384)	-	3.71	3.25	4.50	3.58	3.95
Telephone Communications (481)	4.63	2.81	2.69	2.94	3.07	3.25
Universities (822)	-	2.96	2.72	3.14	3.10	6.01
Laboratory Apparatus (382)	-	2.69	2.73	3.70	3.22	4.04
Investment Companies (679)	-	2.86	2.85	2.79	2.85	3.09
Motor Vehicles (371)	3.69	2.18	2.46	2.87	3.72	3.98
Aircrafts and parts (372)	5.07	2.24	2.47	3.77	3.43	3.06
Professional Equipment Wholesale (504)	-	4.09	3.05	2.60	-	-
Inorganic Chemicals (281)	3.07	2.31	2.50	3.23	3.71	2.35
Engineer.,Architec.,Survey (871)	-	2.74	2.58	3.14	5.33	-
Household Audio-Video (365)	3.49	2.48	2.10	2.04	2.09	2.89
Plastics (282)	3.48	3.79	2.34	2.22	3.50	4.36
Motion Picture Production (781)	-	3.24	3.24	-	-	-
Management,Consulting,PR (874)	-	2.91	3.11	3.38	4.43	-
Electrical Machinery NEC (369)	-	3.04	2.89	3.61	4.38	3.29
Radio and TV Broadcasting (483)	-	3.49	3.48	5.17	-	-
Cable and TV Services (484)	-	4.08	3.23	-	4.10	-
Special Machinery (355)	-	2.89	3.35	3.82	4.44	-
Business Services (738)	-	4.59	3.59	4.01	3.59	-
Crude Oil and Gas (131)	-	3.39	4.08	4.16	6.22	3.59
Electrical Goods Wholesale (506)	-	2.50	2.44	-	-	-
Naut./Aeronaut. Navigation (381)	-	2.53	2.45	4.19	4.10	-
Electric Services (491)	-	2.97	3.81	4.01	3.71	-
Organic Chemicals (286)	3.08	3.86	4.88	4.58	-	4.00

The HE has a minimum either in the 1990-1993 or in the 1994-1997 period, for both the pooled R&D network and most of the sectoral R&D networks. However, 6 out of 30 sectors (Electronic Components, Communications Equipment, Investment Companies, Professional Equipment Wholesale, Household Audio-Video and Plastics) exhibit a minimum in the 1998-2001 period and one single industry (Organic Chemicals) in the early 1986-1989 period. Our results validate hypothesis 4.1: both the pooled and the sectoral R&D networks are characterised by highly dispersed, skewed and fat-tailed degree distributions, as shown by their first four moments as well as by the non-parametric approach of the Hill estimator. In addition, we conclude that such features are non-monotonic: the higher alliance activity of the “golden age” (i.e. a larger average degree) is associated with even more dispersed, skewed and fat-tailed degree distributions.

4.2 Assortative and Disassortative R&D Networks

Assortativity is a network measure that identifies correlations between the centrality of a node and the centrality of its neighbors. Assortativity can be computed by using any measure of node centrality coming from network theory. However, in this study we use degree correlation, or *average nearest-neighbor connectivity* (Newman, 2002; Pastor-Satorras et al., 2001) as assortativity measure. In assortative networks, i.e. networks with positive degree correlation, nodes tend to be connected to other nodes with similar degree, while in disassortative networks, i.e. networks with negative degree correlation, nodes tend to be connected to nodes with dissimilar degree. Newman (2003) found that technological networks, such as the internet, are disassortative while social networks, such as the network of scientific co-authorships, are assortative. However, R&D networks can be assortative or disassortative, depending on the underlying topology of the network. For instance, König and Tessone (2009) and Ramasco et al. (2004) pointed out that limitations in the number of collaborations a firm is able to maintain can give rise to assortativity, even when networks are centralized,⁶ like R&D networks. This means that the firm-level constraint to maintain a limited number of partners can affect the macro-level assortativity measured in R&D networks. Motivated by this, we state the following hypothesis.

Hypothesis 4.2. *R&D networks exhibit positive degree correlations. Firms tend on average to form alliances with firms having similar centrality.*

To test this hypothesis, we use the assortativity mixing coefficient r proposed by Newman (2002). This quantity, as described by Eq. 3, is the Pearson correlation coefficient of the degrees at both ends of all links in the network. The coefficient r ranges between -1 for a totally disassortative network to 1 for a totally assortative network; a network in which links are formed randomly would exhibit $r = 0$. To evaluate r on our networks we use:

$$r = \frac{4M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i (j_i + k_i)]^2}{2M^{-1} \sum_i (j_i^2 + k_i^2) - [M^{-1} \sum_i (j_i + k_i)]^2}, \quad (3)$$

where j_i, k_i are the degrees of the firms at the ends of the i -th link, with $i = 1, \dots, M$. We compute the assortativity mixing coefficient r on both the pooled and the sectoral R&D sub-networks. The whole observation period is again divided into six sub-periods of 4 years each and all the observations of every firm's degree are pooled within each sub-period. The degree correlation coefficients are then computed on these pooled networks in each sub-period. The results are reported in Table 6.

The pooled R&D network exhibits a positive assortative mixing coefficient during the whole observation period. This means that, on average, high-centrality (low-centrality) firms tend to connect to other high-centrality (low-centrality) firms, confirming the hypothesis 4.2. In addition,

⁶Centralized networks consist of a small number of high-degree nodes connected to a large number of low-degree nodes. They usually have a star-like topology and exhibit disassortativity.

Table 6: Assortativity mixing coefficient in the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). Observations are aggregated in six time sub-periods. Missing values refer to networks with too few observations; all the sectoral R&D sub-networks have a significant number of observations in the “golden age”.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	0.167	0.110	0.119	0.195	0.170	0.035
Computer Software (737)	-0.103	-0.074	-0.067	-0.029	-0.002	-0.105
Pharmaceuticals (283)	0.005	0.172	0.119	-0.049	-0.047	-0.043
R&D, Lab and Testing (873)	-0.024	-0.032	0.011	0.132	0.185	0.025
Computer Hardware (357)	-0.188	-0.179	-0.192	-0.133	-0.103	-0.145
Electronic Components (367)	-0.174	-0.151	-0.194	-0.094	0.023	0.267
Communications Equipment (366)	-0.233	-0.149	-0.147	-0.143	-0.077	-0.312
Medical Supplies (384)	-	-0.165	-0.155	0.106	-0.184	-0.108
Telephone Communications (481)	-0.273	-0.178	-0.097	-0.035	-0.036	-0.279
Universities (822)	-	-0.133	-0.102	0.026	0.152	0.078
Laboratory Apparatus (382)	-	-0.199	-0.134	-0.153	-0.159	0.018
Investment Companies (679)	-0.057	-0.210	-0.193	-0.219	-0.187	-0.182
Motor Vehicles (371)	-0.174	-0.309	-0.099	-0.071	-0.023	0.078
Aircrafts and parts (372)	-0.132	0.054	-0.182	0.035	0.019	0.804
Professional Equipment Wholesale (504)	-	-0.128	-0.066	-0.168	-0.200	-
Inorganic Chemicals (281)	-0.445	-0.228	-0.243	-0.188	-0.146	-0.239
Engineer.,Architec.,Survey (871)	-	-0.275	-0.208	-0.130	-0.116	-0.207
Household Audio-Video (365)	-0.467	-0.368	-0.306	-0.329	-0.287	-0.342
Plastics (282)	-0.105	-0.249	-0.351	-0.437	-0.265	-0.151
Motion Picture Production (781)	-	-0.154	-0.081	-0.037	-	-
Management,Consulting,PR (874)	-	-0.288	-0.200	-0.221	-0.177	-0.135
Electrical Machinery NEC (369)	-	-0.250	-0.184	-0.283	-0.032	-0.134
Radio and TV Broadcasting (483)	-	-0.537	-0.173	-0.266	-0.250	-0.250
Cable and TV Services (484)	-	0.006	-0.101	-0.063	-0.287	-
Special Machinery (355)	-	-0.206	-0.223	-0.153	-0.214	-0.143
Business Services (738)	-	-0.296	-0.247	0.382	0.087	-0.100
Crude Oil and Gas (131)	-	0.489	-0.017	0.383	0.255	-0.160
Electrical Goods Wholesale (506)	-	-	-0.305	-0.139	-0.100	-0.143
Naut./Aeronaut. Navigation (381)	-	-0.297	-0.318	-0.333	-0.217	-0.190
Electric Services (491)	-	-0.007	-0.235	-0.127	-0.107	0.664
Organic Chemicals (286)	-0.458	-0.242	-0.206	-0.191	-0.190	-0.170

the non-monotonic trend we observed for other network indicators in Sec. 3 is here replaced by strong fluctuations that do not suggest any periodicity, neither in the pooled nor in the sectoral R&D networks. As opposed to the pooled R&D network, the sectoral R&D networks are mainly disassortative: for most sectors in most of time periods, the assortativity coefficient is indeed negative. For instance, when considering the 1990-1993 and the 1994-1997 periods, only 4 sectors out of 30 exhibit a non-negative assortativity coefficient (Pharmaceuticals, R&D-Lab-Testing, Aircrafts and Parts, Cable and TV Services).

Thus, when taking into account a sectoral R&D network centered around a given industry, low-degree firms increase their tendency to connect to high-degree firms, and viceversa. From a network topology point of view, this is explained by the removal of a non-negligible portion of links when a sectoral network is extracted from the pooled R&D network. What precisely

happens is that the hubs⁷ of each sectoral network (typically belonging to the focal sector) are connected to peripheral nodes (generally not belonging to the focal sector) having a low number of collaborations. We should note here that some of these low-degree nodes could be high-degree nodes in a different sectoral sub-network, or in the pooled R&D network. The fact that in a specific sectoral representation these nodes appear as peripheral nodes is due to the removal of links that are not part of the sectoral representation, as described above.

At the firm level, the above finding implies that when firms have only information about sector-related partnerships (as captured by the sectoral representation of the R&D networks), they tend to form alliances with firms having dissimilar degree. As pointed out by Newman (2002), disassortativity is a typical feature of technological networks. What we observe in the sectoral R&D networks might be that, searching for complementary resources, low-degree firms prefer to ally with high-degree firms, and viceversa. This causes the emergence of disassortativity at the sectoral network level.

Finally, we studied local degree correlations in the pooled R&D network. In Fig. 3 we show the average neighbors' degree as a function of firms' degree, for the pooled R&D network, in six time sub-periods. The observations of every company's degree and average partners' degree are pooled within each sub-period. Each of these six plots can thus be associated with the corresponding assortativity coefficient reported in Table 6 for the pooled R&D network, throughout the six time periods.

We find that the local degree correlation curve is characterized by an inverted U-shaped trend, during the whole observation period. As the slope of the curve testifies, low-degree firms tend to have positive degree correlation, whilst high-degree firms have negative degree correlation. Our finding is that the position of the maximum of these curves on the x -axis (i.e. the firm's degree) varies during the observation period and is positively correlated to the network size. Indeed, such tipping point in the firm's degree is equal to 5 in the early period (1986-1989) and in the late period (2006-2009), and it ranges between 10 and 20 in the other periods. However, during the entire observation period, firms with more than a certain number of partners exhibit a tendency not to form alliances with other high-degree firms. We argue that the reason of this is the finite number of existing high-degree firms – which are, most likely, the biggest players in the various industrial sectors. In addition, the competition effects among these big players might prevent them from forming R&D alliances with one another, giving rise to the negative degree correlation we have unveiled. Finally, we find that the inverted U-shaped trend of the local degree correlation curve holds for the sectoral R&D networks as well. The decreasing part of this curve after the tipping point in the firm's degree is actually more visible in the sectoral R&D networks.⁸

⁷The hubs of a network are the high-degree nodes, i.e. the firms with many partners in our R&D network representation.

⁸We do not show here the local degree correlations for the sectoral R&D networks, but data and plots are available upon request from the authors.

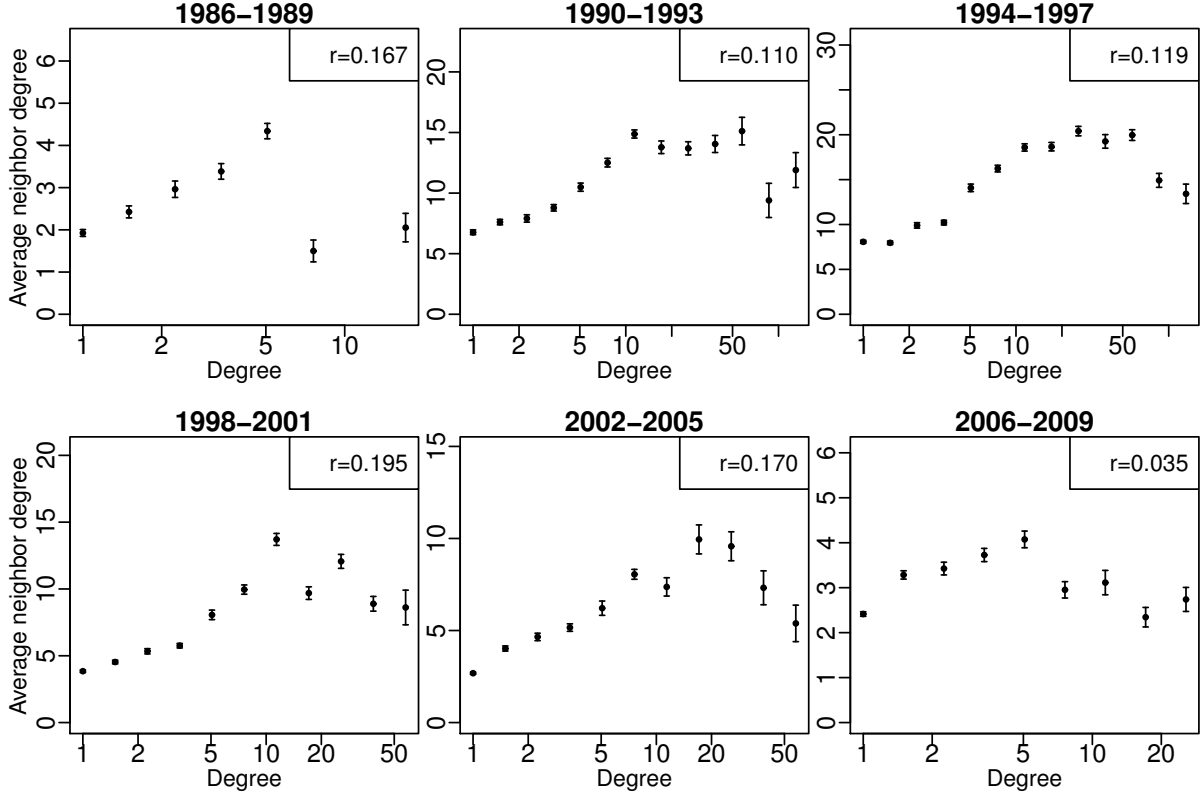


Figure 3: Local degree correlations (mean neighbors' degree VS degree) in the pooled R&D network. The error bars represent the standard error of the mean. On the top-right corner of each plot we report the corresponding value of the assortativity mixing coefficient in the considered time period.

We conclude that the hypothesis 4.2 is validated only for the pooled R&D network, while for the sectoral R&D networks, assortativity is replaced by disassortativity. However, both in the pooled and in the sectoral R&D networks, low-degree firms tend to have positive degree correlation, whilst high-degree firms tend to have negative degree correlation.

5 Small Worlds and Communities

Similarly to degree heterogeneity, the presence of *small worlds* in R&D networks has been analyzed by a large amount of empirical works. A network is a small world if it is characterized by two key features: high local clustering and low average path length, as proposed by Watts and Strogatz (1998). The local clustering measure the extent to which the neighbors of a node are in their turn connected among themselves. It is defined as the number of existing links between the neighbors of a focal node, divided by the number of all possible links between these neighbors; the measure is subsequently averaged over all nodes in the network. The average path length

is defined as the average of all shortest distances, i.e. the lowest number of links that must be traversed to connect every pair of nodes in the network. In our R&D network representation, the first measure shows the extent to which a company's partners know each other, while the second measure quantifies how long the alliance chain from a firm to any other firm in the network is on average. Small world networks exhibit high clustering and short average path length, combining the qualities of both regular networks (typically characterized by high clustering and high average path length) and random networks (characterized by low clustering and low average path length). Previous empirical works have pointed out that the R&D network structure follows a periodic pattern (Gulati et al., 2012; Powell et al., 2005b), in which the excessive formation of ties can lead to the formation of a small world and then to its own decline.

Small world properties in a network are often associated with the presence of community structures (Newman, 2004b), reflecting the tendency of nodes to divide into groups or modules. In a modular network, dense connections and high clustering are observed within each group, with only a few links connecting the different groups (Newman, 2004a). In inter-firm networks, dense groups are shown to facilitate information exchange among similar firms and support trust and cooperative behavior, while bridging ties connecting different groups favour information recombination between distant positions in the knowledge space (e.g. Granovetter, 1973, 1983; Tiwana, 2008).

In this section we analyze both the presence of community structures and of small worlds in R&D networks. In order to assess whether and under what conditions R&D networks can be characterized as small worlds, composed of multiple communities, we state the following hypothesis.

Hypothesis 5.1. *The small world properties of R&D networks are characterized by a rise-and-fall trend and are associated to a modular structure of the R&D network.*

According to Watts and Strogatz (1998), the small world properties of a network have to be evaluated using a corresponding random network as the baseline. If the examined network is both large and sparse, i.e. $N \gg \langle k \rangle$ where N is the network size and $\langle k \rangle$ is the average degree, the basic requirement for small world is satisfied. Under this assumption, the values of clustering coefficient C and average path length L for the baseline random network will tend to: $C_R = \langle k \rangle / N$ and $L_R = \ln(N) / \ln(\langle k \rangle)$. The small world quotient Q_{SW} we use for our analysis is defined as:

$$Q_{SW} = \frac{(C/C_R)}{(L/L_R)}. \quad (4)$$

In our study, the condition of sparse network is always fulfilled for the pooled and the sectoral R&D networks (the average degree is always smaller than three, as reported in Table 4). Some of the sectoral R&D networks do not have a large size in the first (1986-1989) and in the last (2006-2009) observation periods, as can be seen from Table 1, but in these cases they exhibit an even smaller average degree $\langle k \rangle$. When computing the observed to random ratios, a small

world network will show $C/C_R \gg 1$ and $L/L_R \simeq 1$, which is the case for almost all the R&D networks we analyse. Our results are listed in Table 7.

Table 7: Small world quotient for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). The values are averaged in six time sub-periods. Missing values refer to networks with too few observations; all the sectoral R&D sub-networks have a significant number of observations in the “golden age”.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	12.90	116.76	200.34	84.9	57.75	26.46
Computer Software (737)	4.58	18.98	45.08	9.2	4.77	0.00
Pharmaceuticals (283)	1.99	29.96	41.03	20.9	9.26	5.39
R&D, Lab and Testing (873)	0.00	7.25	18.81	8.0	7.31	4.51
Computer Hardware (357)	0.16	5.99	21.29	9.5	1.09	0.00
Electronic Components (367)	2.43	9.87	16.36	10.1	9.69	7.89
Communications Equipment (366)	0.00	3.36	7.94	3.2	3.92	0.65
Medical Supplies (384)	-	2.38	1.59	6.0	0.99	3.05
Telephone Communications (481)	0.00	10.91	13.76	2.8	4.46	1.16
Universities (822)	0.00	3.56	6.55	4.9	7.66	11.08
Laboratory Apparatus (382)	-	1.35	0.67	1.2	0.00	2.75
Investment Companies (679)	0.00	2.68	3.96	2.7	3.85	2.46
Motor Vehicles (371)	2.32	4.52	6.48	1.9	4.55	4.39
Aircrafts and parts (372)	2.11	6.29	7.39	5.4	8.01	8.12
Professional Equipment Wholesale (504)	0.00	4.56	2.85	0.0	0.00	-
Inorganic Chemicals (281)	0.00	0.55	1.10	0.0	0.00	0.00
Engineer.,Architec.,Survey (871)	-	2.04	1.65	0.0	0.00	0.00
Household Audio-Video (365)	0.00	2.83	7.60	4.7	2.55	1.92
Plastics (282)	0.00	0.47	0.33	0.0	0.00	0.00
Motion Picture Production (781)	-	0.00	0.00	0.0	-	-
Management,Consulting,PR (874)	-	0.00	0.00	0.0	0.00	0.00
Electrical Machinery NEC (369)	-	0.22	2.46	0.0	0.00	0.00
Radio and TV Broadcasting (483)	0.00	2.06	1.55	0.0	0.00	0.00
Cable and TV Services (484)	-	0.69	5.19	0.0	0.00	-
Special Machinery (355)	-	0.92	1.14	0.0	0.00	0.00
Business Services (738)	-	0.00	0.00	8.8	4.43	0.00
Crude Oil and Gas (131)	0.00	7.70	5.29	4.5	10.83	0.00
Electrical Goods Wholesale (506)	-	0.00	0.00	0.0	0.00	0.00
Naut./Aeronaut. Navigation (381)	0.00	0.00	0.00	0.0	0.00	0.00
Electric Services (491)	-	4.88	1.03	3.2	2.23	11.10
Organic Chemicals (286)	0.00	0.00	0.00	0.0	0.00	0.00

The small world quotient has been computed separately for every year during the whole observation period, in both the pooled and the sectoral R&D networks, and then averaged within six time sub-periods lasting 4 years each. We did not pool the observations inside every time period, because the small world quotient is a pooled network measure, and not an ego-network measure centered around single nodes. The trend of this indicator supports our hypothesis of non-monotonic small world properties, in both the pooled R&D network and the sectoral R&D networks. The small world quotient rises to a peak in the “golden age” in the first half of the nineties and then decreases again. Moreover, this feature is common across sectors, extending the results of Gulati et al. (2012), who found this inverse U-shaped trend in the computer industry. We argue that such non-monotonic trend is the first evidence of the rise-and-fall of R&D networks; it is the empirical confirmation of the emergence of a complex structure (namely,

small world) in the R&D networks during the “golden age”. With the exception of 6 sectors out of 30 (Medical Supplies, Universities, Aircrafts and Parts, Business Services, Crude Oil and Gas, Electric Services), the small world quotient has a peak either in the 1990-1993 or in the 1994-1997 period. It should also be noticed that five industrial sectors (Motion Picture Production, Management-Consulting-P.R., Electrical Goods Wholesale, Nautical/Aeronautical Navigation, Organic Chemicals) display constant zero values for their small world quotients, meaning that in the correspondent networks there is no observed clustering. The sectors that deviate the most from the non-monotonic small world trend are mostly service sectors, which indeed tend to create more inter-sectoral alliances, rather than forming their own intra-sectoral network.

We now want to assess whether such emergence of small world properties in R&D networks is associated with the presence of modular structures. The standard approach to quantify this, described by Newman (2004b), is to perform a partition of the network into communities, i.e. assigning a label to every node, in order to maximize the so called *modularity coefficient*. Such indicator of modularity is maximum if the chosen network partition perfectly reflects the positioning of links in the network, with all links occurring within communities and no links occurring between different communities. Following this approach, a community corresponds to an industrial sector in our R&D network representation. We consider the pooled R&D network and partition it by assigning every firm to its sector. However, we do not test several partitions to maximize the modularity coefficient; we rather take the partition of the pooled R&D network into communities as fixed and study the *evolution* over time of the modularity coefficient computed on top of this community structure. We call the modularity coefficient Q_M and define the *relative connectivity* c_{ij} between two industrial sectors i and j as follows:

$$c_{ij} = e_{ij}/a_{ij}, \quad (5)$$

where e_{ij} is the fraction of links in the network connecting any firm belonging to sector i to any firm belonging to sector j . The quantities e_{ij} , a_{ij} , and consequently c_{ij} , can be thought of as elements of three symmetric $n \times n$ matrices, where n is the number of sectors into which the R&D network has been partitioned.⁹ The row (or column) sums $a_i = \sum_j e_{ij}$ represent the fraction of links (alliances) involving at least one company in sector i . This way, $a_{ij} = a_i a_j$ is the expected fraction of links connecting firms in sector i to firms in sector j in a benchmark network having the same density and sector populations as the real network, but where alliances occur randomly between firms, independently of the sector they belong to. It is clear that c_{ij} indicates the ratio between the observed and the expected fraction of alliances connecting a firm in sector i to a firm in sector j . Values of c_{ij} greater than 1 suggest that the alliance probability between a firm in sector i and a firm in sector j is higher than one would expect with a random partner choice. On the contrary, when c_{ij} is smaller than 1, a firm in sector i forms alliances with firms in sector j with a smaller probability than a random partner choice. The relative connectivities c_{ij} between

⁹To make sure that every alliance is counted once in the matrix e_{ij} , every link connecting sectors i and j is split in half between the elements e_{ij} and e_{ji} .

the 18 largest industrial sectors we selected for our study are listed in Table 8; the pooled R&D network in the year 1995 has been used to compute these values.

Table 8: Relative connectivities between the 18 largest sectors.

	283	737	873	367	357	384	366	679	481	822	382	371	874	281	372	871	131	504
Pharmaceuticals (283)	6.5	0.1	3.1	0.1	0.1	2.2	0.2	1.0	0.1	3.6	1.6	0.2	0.4	2.4	0.0	0.0	0.0	0.8
Computer Software (737)	0.1	3.4	0.5	1.3	2.2	0.2	1.4	1.4	1.5	0.6	1.3	0.5	3.0	0.6	0.6	1.0	0.1	2.2
R&D, Lab and Testing (873)	3.1	0.5	7.7	0.3	0.2	2.1	0.2	2.1	0.9	4.0	2.8	0.9	4.3	2.3	0.0	2.3	0.0	1.7
Electronic Components (367)	0.1	1.3	0.3	5.0	3.2	1.0	3.3	1.7	0.8	0.7	1.3	0.0	0.8	0.6	0.0	3.3	0.0	1.8
Computer Hardware (357)	0.1	2.2	0.2	3.2	3.5	0.4	2.2	0.9	1.2	0.4	0.5	0.5	1.1	0.4	0.2	0.4	0.0	3.0
Medical Supplies (384)	2.2	0.2	2.1	1.0	0.4	29.4	0.2	1.8	0.0	1.3	2.7	0.8	0.0	0.0	0.0	0.0	0.0	0.0
Communications Equipment (366)	0.2	1.4	0.2	3.3	2.2	0.2	7.2	1.4	2.4	0.2	2.1	0.4	0.0	0.0	0.9	1.8	0.0	2.0
Investment Companies (679)	1.0	1.4	2.1	1.7	0.9	1.8	1.4	9.1	1.0	1.1	2.1	2.4	2.2	3.1	0.8	6.2	3.8	0.0
Telephone Communications (481)	0.1	1.5	0.9	0.8	1.2	0.0	2.4	1.0	14.9	0.8	1.0	0.4	4.7	0.0	0.6	0.0	2.0	0.0
Universities (822)	3.6	0.6	4.0	0.7	0.4	1.3	0.2	1.1	0.8	6.9	3.5	4.4	0.0	1.9	1.9	0.0	3.5	1.1
Laboratory Apparatus (382)	1.6	1.3	2.8	1.3	0.5	2.7	2.1	2.1	1.0	3.5	10.7	0.0	0.0	3.9	7.8	0.0	0.0	0.0
Motor Vehicles (371)	0.2	0.5	0.9	0.0	0.5	0.8	0.4	2.4	0.4	4.4	0.0	60.8	0.0	0.0	4.9	13.1	2.0	0.0
Management, Consulting, PR (874)	0.4	3.0	4.3	0.8	1.1	0.0	0.0	2.2	4.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Inorganic Chemicals (281)	2.4	0.6	2.3	0.6	0.4	0.0	0.0	3.1	0.0	1.9	3.9	0.0	0.0	85.3	4.3	0.0	5.1	0.0
Aircrafts and parts (372)	0.0	0.6	0.0	0.0	0.2	0.0	0.9	0.8	0.6	1.9	7.8	4.9	0.0	4.3	76.7	25.6	2.6	0.0
Engineer.,Architec.,Survey (871)	0.0	1.0	2.3	3.3	0.4	0.0	1.8	6.2	0.0	0.0	0.0	13.1	0.0	0.0	25.6	0.0	5.1	0.0
Crude Oil and Gas (131)	0.0	0.1	0.0	0.0	0.0	0.0	0.0	3.8	2.0	3.5	0.0	2.0	0.0	5.1	2.6	5.1	118.1	0.0
Profess. Equipment Wholesale (504)	0.8	2.2	1.7	1.8	3.0	0.0	2.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.4

As reported in Table 8, most of the connectivity values in the main diagonal c_{ii} , that we define as the *intra-sector connectivities*, are greater than 1. On the contrary, the elements c_{ij} outside the main diagonal, that we define as the *inter-sector connectivities*, are instead smaller than 1, or anyway smaller than the corresponding diagonal element. In other words, most of the alliances we have tracked in the pooled R&D network occur within sectors rather than between different sectors. There are only two exceptions: Management-Consulting-PR and Engineering-Architecture-Survey, that – being service sectors – have a natural bias towards inter-sectoral alliances, rather than intra-sectoral alliances. Such cross-sector investigation confirms that the choice of industrial sectors as network communities reflects quite well the alliance activity of firms. Given this community structure as fixed, we define the modularity coefficient Q_M as:

$$Q_M = \sum_i (e_{ii} - a_{ii}^2) / (1 - \sum_i a_{ii}^2), \quad (6)$$

where, following the previous definitions, the index i spans all industrial sectors in the R&D network. The coefficient Q_M is equal to 1 in case of a perfect modular network, where alliances occur only intra-community and never inter-community. Likewise, Q_M is equal to -1 for a perfect anti-modular network, having only inter-community links, without any intra-community links. Q_M is equal to zero for a network where links are formed at random. We compute the modularity coefficient Q_M on the pooled as well as the sectoral R&D networks and report its evolution over time in Table 9.

The coefficient Q_M of the pooled R&D network ranges between 0.21 and 0.28, showing a moderate modularity if compared to other examples of real networks (see Newman and Girvan, 2004). However, the modularity coefficient exhibits only small changes over the observation period and does not have a peak in accordance with the peak of the small world quotient. The rise-and-fall of the small world structure is thus not simultaneous to any rise-and-fall in the modular structure, which is instead a constant feature of the pooled R&D network.

Table 9: Modularity coefficients for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). The values are averaged in six time sub-periods. Missing values refer to networks with too few observations; all the sectoral R&D sub-networks have a significant number of observations in the “golden age”.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	0.237	0.220	0.228	0.220	0.218	0.277
Computer Software (737)	-0.163	-0.134	-0.058	-0.112	-0.178	-0.241
Pharmaceuticals (283)	0.107	-0.009	-0.048	-0.010	-0.091	-0.024
R&D, Lab and Testing (873)	-0.196	-0.207	-0.176	-0.171	-0.270	-0.260
Computer Hardware (357)	-0.532	-0.251	-0.234	-0.265	-0.259	-0.217
Electronic Components (367)	-0.023	-0.145	-0.210	-0.211	-0.180	-0.083
Communications Equipment (366)	-0.250	-0.159	-0.215	-0.208	-0.218	-0.243
Medical Supplies (384)	-0.174	-0.206	-0.165	-0.255	-0.283	-0.282
Telephone Communications (481)	-0.185	-0.021	-0.120	-0.229	-0.192	-0.259
Universities (822)	-0.721	-0.280	-0.268	-0.236	-0.250	-0.379
Laboratory Apparatus (382)	-0.203	-0.155	-0.176	-0.202	-0.194	-0.267
Investment Companies (679)	-0.390	-0.195	-0.151	-0.124	-0.153	-0.193
Motor Vehicles (371)	0.235	-0.041	0.048	0.029	0.111	0.052
Aircrafts and parts (372)	-0.079	0.000	0.016	0.069	-0.053	0.063
Professional Equipment Wholesale (504)	-0.761	-0.330	-0.275	-0.286	-0.306	-0.356
Inorganic Chemicals (281)	-0.192	-0.159	-0.108	-0.128	-0.170	-0.182
Engineer.,Architec.,Survey (871)	0.000	-0.140	-0.135	-0.083	-0.140	-0.213
Household Audio-Video (365)	-0.290	-0.227	-0.202	-0.217	-0.259	-0.264
Plastics (282)	-0.312	-0.212	-0.172	-0.201	-0.226	-0.191
Motion Picture Production (781)	-	0.000	0.000	0.000	0.000	0.000
Management,Consulting,PR (874)	0.000	-0.244	-0.169	-0.135	-0.232	-0.170
Electrical Machinery NEC (369)	0.000	-0.187	-0.133	-0.217	-0.154	-0.105
Radio and TV Broadcasting (483)	0.000	0.000	0.000	0.000	0.000	0.000
Cable and TV Services (484)	-	0.000	0.000	0.000	0.000	0.000
Special Machinery (355)	-0.536	-0.231	-0.198	-0.067	-0.031	-0.155
Business Services (738)	-1.000	-0.516	-0.206	-0.177	-0.174	-0.573
Crude Oil and Gas (131)	-0.568	0.064	0.052	0.067	-0.049	-0.096
Electrical Goods Wholesale (506)	-	0.000	0.000	0.000	0.000	0.000
Naut./Aeronaut. Navigation (381)	-0.600	-0.257	-0.226	-0.132	-0.214	-0.272
Electric Services (491)	-	-0.058	-0.091	-0.135	-0.087	-0.171
Organic Chemicals (286)	-0.437	-0.205	-0.177	-0.192	-0.278	-0.266

We computed the modularity coefficient for the sectoral networks as well. Our finding is that they mostly exhibit negative values of Q_M . We could anticipate this result, as a sectoral R&D network includes all the alliances between firms in the focal sector and all other sectors, but no alliances between firms simultaneously not belonging to the focal sector. This destroys all community structures outside the focal sector, leaving a disproportionate number of alliances linking the focal sector with the non-focal ones. For this reason, the resulting modularity coefficient has negative values for most of the sectors in most of the time periods. Four sectors show a modularity coefficient constantly equal to 0, indicating no preferences at all in forming inter- or intra-sectoral alliances: Motion Picture Production, Radio and TV Broadcasting, Cable and TV Services and Electrical Goods Wholesale. Overall, the trend of this indicator in the sectoral R&D networks exhibits a certain degree of fluctuation, without the presence of any non-monotonic trend.

In conclusion, we can partly validate hypothesis 5.1, stating that both the pooled and the

sectoral R&D networks exhibit the rise-and-fall of small world properties. However, this feature is not associated, neither at the pooled nor at the sectoral level, with the presence of a strong community division in the network. The emergence of small world properties might thus have other reasons, which will be further investigated in the next section.

6 Core-Periphery and Nested Architectures

In general, the presence of hierarchical structures is a determinant factor for the behavior of a network. This is because hierarchical networks are typically strongly centralized, i.e. dominated by a group of highly inter-connected nodes (the *core* of the network) having a few connections to secondary nodes (the *periphery* of the network). Borgatti (2005) points out that strongly centralized networks are efficient because they can spread information quickly. Moreover, König et al. (2012) show that efficient R&D network structures have a *nested* architecture, when marginal costs of collaborations are high. Previously used as a measure of order in ecological systems (see Bascompte et al., 2003), *nestedness* is considered here in terms of nested neighborhood structures: a network is said to be nested if the neighborhood of a node is contained in the neighborhoods of nodes with higher degree. In addition, nestedness can be thought of as a refinement of the concept of core-periphery. A core-periphery architecture features a core of densely connected nodes and a periphery of nodes that are linked to the core, but only sparsely inter-connected among themselves (Goyal, 2007). Nested networks can have not only two, but several inter-connected groups of nodes with increasing degree, each of them being connected to the group of higher-degree nodes.

Finally, both core-periphery and nested networks can exhibit short path length and high clustering (König et al., 2009), features that are typical for small worlds. In our case, given the absence of correlation between the emergence of small worlds and modular architectures in the R&D networks, the formation of hierarchical structures could be the true reason for the emergence of small world properties that were reported in Sec. 5. Motivated by this reasoning, and in order to evaluate whether the small world properties of the R&D networks are associated with the emergence of hierarchical structures, we state the following hypothesis.

Hypothesis 6.1. *Hierarchical architectures, namely core-periphery and nested structures, show a non-monotonic trend in accordance to the emergence of small world properties in the R&D networks.*

The measure we use to quantify the presence of a core-periphery architecture in our R&D networks is a slightly modified version of the core-periphery coefficient C_{cp} suggested by Holme (2005). As opposed to Holme (2005), who suggests to calculate such indicator only on the largest connected component of the network, we take into account the whole network. We then define

the core-periphery coefficient C_{cp} of a network G as follows:

$$C_{cp} = \frac{c_c[G^{core}]/c_c[G]}{c_c[G_R^{core}]/c_c[G_R]}, \quad (7)$$

where $c_c[\cdot]$ indicates the closeness centrality of a network¹⁰ and G^{core} is a subgraph¹¹ of the network G that maximises this value of closeness centrality. This ratio is computed on the examined network and then divided by the same measure computed as the mean value over 500 random networks of the same size and density as the examined network G . The values of the core-periphery coefficients C_{cp} for the pooled and the sectoral R&D networks are shown in Table 10. As the core-periphery coefficient is a global network measure, we did not pool the observations inside each of the 6 selected time periods, but we computed its value separately for every year and then averaged over the duration of every time period.

We clearly observe a rise-and-fall trend for the core-periphery coefficient, in both the pooled and the sectoral R&D networks, with a peak positioned either in the 1990-1993 or in the 1994-1997 period. The presence of this core-periphery structure associated with small-worldliness in the “golden age” is a common characteristic across all industrial sectors. One notable exception is the Pharmaceutical sector, whose core-periphery properties obtain their highest values in the period 2002-2005. In addition, four small industrial sectors (Management-Consulting-PR, Business Services, Electrical Goods Wholesale and Organic Chemicals) exhibit core-periphery coefficients that are not peaked neither in 1990-1993 nor in the 1994-1997 periods.

This means that – both at pooled and sectoral level – R&D networks are characterised by small world features that are associated with the emergence of a strong centralization, rather than modularity. Across sectors, firms show the tendency to organise their R&D collaborations in a core of densely connected companies and a periphery of companies that are linked to the core, but only weakly interconnected among themselves.

Next, we study whether the R&D networks are specifically characterised by nested neighborhood structures. A network is defined to be nested if the neighbors of a node with degree m are contained in the neighborhoods of all nodes with degree $m' > m$. There exist a number of measures quantifying the extent to which a given network’s neighborhood structure is nested. In this study, we use the *BINMATNEST* algorithm,¹² that returns nestedness coefficients C_n ranging from 0 (for a totally nested network) to 100 (for a completely random, non-nested network).

¹⁰The closeness centrality of a network is defined as the inverse of the sum of all shortest paths between any pair of nodes in the network. The idea behind this measure is to quantify how connected a network is. See Sabidussi (1966) for a more rigorous definition.

¹¹There are many ways to divide a network G into subgraphs and then select the subgraph G^{core} with the maximal closeness centrality. Usually, one uses the computationally cheapest algorithm, which is a k -core decomposition of the network. G^{core} is then assumed to be the k -shell of the network with maximal closeness centrality. For the sake of brevity, we do not provide here any description of the k -core decomposition procedure see Sabidussi (1966) for a detailed explanation, and Garas et al. (2012) for an extension to weighted networks.

¹²The *BINMATNEST* algorithm has been proposed by Rodriguez-Girones and Santamaria (2006).

Table 10: Core-periphery coefficients for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). The values are averaged in six time sub-periods. Missing values refer to networks with too few observations; all the sectoral R&D sub-networks have a significant number of observations in the “golden age”.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	8.37	23.53	28.51	20.13	18.46	16.34
Computer Software (737)	4.62	8.48	16.38	4.34	5.44	0.03
Pharmaceuticals (283)	0.97	11.41	7.59	12.69	12.88	3.81
R&D, Lab and Testing (873)	0.16	4.01	11.36	10.69	5.47	0.49
Computer Hardware (357)	0.80	5.04	12.38	8.43	2.33	0.12
Electronic Components (367)	1.20	7.17	11.63	6.87	4.39	2.49
Communications Equipment (366)	0.19	3.13	11.05	4.11	2.30	0.77
Medical Supplies (384)	0.30	0.86	3.25	0.98	1.67	1.93
Telephone Communications (481)	0.39	10.85	14.21	3.12	0.96	0.70
Universities (822)	0.81	0.95	8.47	2.35	1.93	2.35
Laboratory Apparatus (382)	0.34	3.95	3.62	2.57	0.04	2.27
Investment Companies (679)	0.26	5.75	7.84	7.68	4.91	2.33
Motor Vehicles (371)	0.69	3.25	8.48	2.20	0.91	0.36
Aircrafts and parts (372)	1.87	5.75	9.12	3.18	1.01	1.88
Professional Equipment Wholesale (504)	0.58	1.94	2.75	0.15	0.38	0.35
Inorganic Chemicals (281)	0.17	3.60	1.89	0.06	0.07	0.09
Engineer.,Architec.,Survey (871)	1.00	3.12	2.85	0.05	0.15	0.19
Household Audio-Video (365)	0.50	2.98	10.12	5.43	2.96	1.95
Plastics (282)	0.28	2.37	3.80	0.07	0.08	0.16
Motion Picture Production (781)	-	0.78	1.45	0.30	0.66	0.55
Management,Consulting,PR (874)	1.00	0.19	0.03	0.05	0.05	0.12
Electrical Machinery NEC (369)	1.00	2.12	4.24	0.10	0.14	0.09
Radio and TV Broadcasting (483)	1.00	2.40	2.62	0.13	0.45	0.64
Cable and TV Services (484)	-	0.81	4.43	0.36	0.48	0.78
Special Machinery (355)	0.89	0.93	1.60	0.08	0.24	0.28
Business Services (738)	1.00	0.46	0.04	1.25	0.72	0.67
Crude Oil and Gas (131)	0.65	3.45	4.20	2.59	1.03	0.11
Electrical Goods Wholesale (506)	-	0.23	0.04	0.18	0.29	0.38
Naut./Aeronaut. Navigation (381)	1.00	0.78	1.20	0.14	0.19	0.24
Electric Services (491)	-	3.51	1.37	1.42	1.37	2.27
Organic Chemicals (286)	0.63	0.11	0.06	0.15	0.13	0.16

In order to have a benchmark for every analysed network, the algorithm builds 500 corresponding random networks with the same size and density, and computes their nestedness coefficients. We use a normalised nestedness coefficient C'_n , defined as:

$$C'_n = \frac{100 - C_n}{100}, \quad (8)$$

where C_n is the nestedness coefficient generated by the *BINMATNEST* algorithm. The normalised nestedness coefficient C'_n spans thus from 0, for a for a totally non-nested network, to 1, for a totally nested network. We calculated the coefficients C'_n throughout the whole observation period, for the pooled and the sectoral R&D networks, and we averaged the results within six time sub-periods lasting 4 years each. Results are shown in Table 11.

The values of the nestedness coefficients C'_n we report are close to 1, during the whole observation period, both for the pooled and the sectoral R&D networks. This is surprising, if we compare such values with other studies of nestedness in real networks (e.g. Bascompte et al.,

Table 11: Nestedness coefficients for the pooled and the sectoral R&D networks (SIC codes are indicated in parentheses). The values are averaged in six time sub-periods. Missing values refer to networks with too few observations; all the sectoral R&D sub-networks have a significant number of observations in the “golden age”.

Sector	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled	0.977	0.997	0.999	0.997	0.996	0.996
Computer Software (737)	0.981	0.992	0.997	0.985	0.950	0.946
Pharmaceuticals (283)	0.960	0.989	0.996	0.994	0.995	0.996
R&D, Lab and Testing (873)	0.961	0.969	0.992	0.986	0.986	0.975
Computer Hardware (357)	0.960	0.992	0.995	0.984	0.950	0.940
Electronic Components (367)	0.981	0.984	0.993	0.984	0.969	0.926
Communications Equipment (366)	0.943	0.962	0.990	0.973	0.944	0.962
Medical Supplies (384)	0.998	0.944	0.963	0.954	0.946	0.947
Telephone Communications (481)	0.945	0.954	0.981	0.950	0.947	0.976
Universities (822)	0.961	0.971	0.973	0.952	0.948	0.958
Laboratory Apparatus (382)	0.961	0.943	0.965	0.930	0.951	0.924
Investment Companies (679)	0.956	0.973	0.979	0.961	0.962	0.940
Motor Vehicles (371)	0.938	0.961	0.964	0.950	0.962	0.946
Aircrafts and parts (372)	0.945	0.942	0.969	0.973	0.953	0.974
Professional Equipment Wholesale (504)	0.911	0.930	0.930	0.952	0.921	0.998
Inorganic Chemicals (281)	0.930	0.978	0.951	0.951	0.943	0.977
Engineer.,Architec.,Survey (871)	-	0.924	0.917	0.962	0.957	0.998
Household Audio-Video (365)	0.951	0.945	0.981	0.964	0.957	0.963
Plastics (282)	0.977	0.940	0.966	0.951	0.975	0.949
Motion Picture Production (781)	-	0.935	0.925	0.923	0.937	-
Management,Consulting,PR (874)	-	0.932	0.933	0.954	0.959	0.939
Electrical Machinery NEC (369)	0.939	0.947	0.950	0.962	0.961	0.987
Radio and TV Broadcasting (483)	0.939	0.941	0.969	0.942	0.967	0.958
Cable and TV Services (484)	-	0.930	0.925	0.975	0.973	-
Special Machinery (355)	-	0.927	0.940	0.984	0.931	0.953
Business Services (738)	0.901	0.926	0.952	0.954	0.972	0.998
Crude Oil and Gas (131)	0.939	0.922	0.950	0.945	0.960	0.936
Electrical Goods Wholesale (506)	-	0.951	0.940	0.953	0.935	0.998
Naut./Aeronaut. Navigation (381)	0.939	0.938	0.948	0.936	0.982	0.998
Electric Services (491)	0.956	0.918	0.969	0.977	0.957	0.939
Organic Chemicals (286)	0.956	0.938	0.961	0.922	0.939	0.956

2003). All the values found in our R&D networks are significantly different from the average values of the random networks used as benchmark in the *BINMATNEST* algorithm. Moreover, the level of nestedness has a peak during the “golden age” for almost the totality of the sectors. The emergence of nested structures in the R&D networks supports the prediction of the model described by König et al. (2012), that studies the efficiency and stability of R&D networks with network-dependent direct and indirect spillovers. In this model, the efficient network structure is shown to critically depend on the marginal cost of R&D collaborations. In case of relatively costly partnerships, the resulting efficient R&D network exhibits a strongly nested neighborhood structure, as we observe empirically.

In conclusion, we validate the hypothesis 6.1, with regard to the rise-and-fall of core-periphery and nested structures, both for the pooled and the sectoral R&D networks. The emergence of core-periphery and nestedness properties in the pooled and the sectoral R&D networks exhibits a peak in the “golden age”.

7 Concluding Remarks

In this study we empirically analyzed the evolution of the pooled R&D network and a set of sectoral R&D networks from 1986 to 2009. We showed that the network size peaks in 1995 and then decreases again until the end of the observation period, for both the pooled and the sectoral networks. The fraction of nodes belonging to largest connected component follows the same trend, reaching a peak in the mid-nineties and then shrinking again, leaving space to a periphery of disconnected dyads. The emergence of a giant component is then followed by a non-monotonic change in its size, with a peak in the mid-nineties. In addition, we find that this feature is robust, as we see it both in the pooled and in all sectoral R&D networks we study. Such universal behavior across sectors is also displayed by more sophisticated network measures and leads us to define the years between 1990 and 1997 as the “golden age” of R&D networks.

Furthermore, we show that R&D networks across sectors are characterized by highly dispersed, skewed and fat-tailed degree distributions, along the whole observation period. The first four moments of the degree distributions are never compatible with a normal distribution, highlighting a huge heterogeneity in firms’ behavior, independent of their industry. In addition, we observe that a greater average degree (occurring in the “golden age”) is associated with even more dispersed, skewed and fat-tailed degree distributions, indicating that higher alliance activity is related to higher heterogeneity. We also show that the pooled R&D network is overall assortative. However, such an outcome is probably the result of a composition effect. Indeed, when looking in detail at the firm level, we observe that assortativity is related to firm degree. More precisely, we found low-degree firms are assortative while high-degree firms are disassortative. In addition, such an inverse U-shaped trend holds across sectors and during the whole observation period. Finally, when shifting the focus from the pooled R&D network to the sectoral R&D sub-networks, the overall assortativity is replaced by disassortativity.

The above patterns support the idea that while searching for complementary resources, low-degree firms prefer to ally with high-degree firms, and vice-versa, giving rise to such negative degree correlation observed at the sectoral level. Moreover, the existence of a limited number of big firms in each sector and competition effects might have prevented those firms from forming many intra-sectoral R&D alliances with one another, thereby giving rise to the disassortativity we have unveiled for high-degree firms.

Furthermore, both the pooled and the sectoral R&D networks are characterized by non-monotonic small world properties as well, with a peak in the mid-nineties. However, this feature is not associated with the presence of a higher modularity in the network. The emergence of small-world properties is instead associated with the presence of core-periphery structures, both in the pooled and in the sectoral R&D networks. More specifically, we have shown that the R&D networks are characterized by strongly nested neighborhood structures, with a non-monotonic trend peaked in the mid-nineties, supporting previous theoretical models on R&D network formation.

To sum up, our analysis shows that many features of R&D networks, namely heterogeneous degree distributions, small world properties, presence of core-periphery and nested structures, are invariant with respect to the scale of aggregation of the R&D networks. The presence of a rise-and-fall trend has been already detected in relation to the small world property of the network (see e.g. Gulati et al., 2012). However, our results show that this trend is a general feature of many – and in general more complex – network properties.

We can draw three fundamental conclusions in this regard. First, we unveiled the emergence of complex structures in R&D networks after an initial phase (until 1989). Namely, such structures are represented by the giant component of the network, small worlds, core-periphery and nested architectures. Second, we have shown that, after their emergence, these structures change their size over time non-monotonically, reaching a peak in the mid-nineties and then decreasing again. Third, this feature is robust, as we found it in almost all the sectoral R&D networks we analyzed. These similarities across sectors lead us to define the years between 1990 and 1997 as the “golden age” witnessing the *rise-and-fall* of R&D networks. We do not observe a second rising part in the trends we study, supposedly because of the time limitation of the dataset we use. We certainly do not exclude a subsequent second rise of the R&D networks, following the phase of change and re-adaptation that firms experience in the late period until 2009. R&D networks might actually be characterized by what we call a *life cycle*.

Furthermore, our analysis unveiled that R&D networks dynamics looks rather independent from technological differences across sectors. Sectoral R&D networks exhibited similar characteristics among themselves. A possible interpretation is that R&D networks dynamics is a highly path-dependent process, and the future creation/destruction of alliances is determined by the current structure of the network. In turn, this is in line with the predictions of theoretical models of strategic network formation (e.g. Goyal and Joshi, 2003; König et al., 2012; König and Tes-sone, 2009; Westbrock, 2010), where the topology of the network and the centrality of individual nodes are relevant to determine which links are formed. In this approach, the change of the technological landscape where firms move can be a consequence of the alliance network in which they are embedded.

This work could be extended at least in two ways. First, one could track the evolution of the creation/destruction of single R&D alliances and test whether this can be predicted based on some topological properties of the existing networks, according to the predictions of theoretical models of strategic network formation. Another extension could involve the study of the mutual feedback between network position and technological position of firms, to unveil the effects of alliances on technology change. We plan to extend our work to empirical data about firms’ patenting, that would allow us to measure the evolution of technology portfolios as a consequence of R&D alliance formation.

8 Acknowledgements

M. V. T, acknowledges financial support from the SNF through grant 100014_126865. A. G. and F. S. acknowledge support by the EU-FET project MULTIPLEX 317532.

References

- Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, 45(3):425–455.
- Barabasi, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.
- Bascompte, J., Jordano, P., Melián, C. J., and Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387.
- Bojanowski, M., Corten, R., and Westbrook, B. (2012). The structure and dynamics of the global network of inter-firm R&D partnerships 1989-2002. *Journal of Technology Transfer*, 37:967–987.
- Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(2):55–71.
- Deeds, D. and Hill, C. (1999). An examination of opportunistic action within research alliances—the analysis of discrete structural alternatives. *Journal of Business Venturing*, 14(2):141–163.
- Fleming, L., King, C., and Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science*, 18(6):938–954.
- Fleming, L. and Marx, M. (2006). Managing creativity in small worlds. *California Management Review*, 48(4):6–27.
- Garas, A., Schweitzer, F., and Havlin, S. (2012). A k -shell decomposition method for weighted networks. *New Journal of Physics*, 14(8):083030.
- Goyal, S. (2007). *Connections: an introduction to the economics of networks*. Princeton University Press. forthcoming.
- Goyal, S. and Joshi, S. (2003). Networks of collaboration in oligopoly. *Games and Economic behavior*, 43(1):57–85.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78:1360–1380.

- Granovetter, M. S. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233.
- Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative science quarterly*, pages 619–652.
- Gulati, R., Sytch, M., and Tatarynowicz, A. (2012). The rise and fall of small worlds: Exploring the dynamics of social structure. *Organization Science*, 23(2):449–471.
- Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research policy*, 31(4):477–492.
- Hanaki, N., Nakajima, R., and Ogura, Y. (2007). The dynamics of R&D collaboration in the IT industry. Working Paper.
- Hanaki, N., Nakajima, R., and Ogura, Y. (2010). The dynamics of R&D network in the it industry. *Research policy*, 39(3):386–399.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):pp. 1163–1174.
- Holme, P. (2005). Core-periphery organization of complex networks. *Phys. Rev. E*, 72:046111.
- Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.
- König, M., Battiston, S., Napoletano, M., and Schweitzer, F. (2012). The efficiency and stability of R&D networks. *Games and Economic Behavior* (forthcoming).
- König, M. D. and Tessone, C. J. (2009). The emergence of hierarchies in networks of agents competing for high centrality. Submitted to Physical Review Letters.
- König, M. D., Tessone, C. J., and Zenou, Y. (2009). Games of dynamic network formation. Submitted to Journal of Economic Theory.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701.
- Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, 67(2):026126.
- Newman, M. E. J. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(90001):5200–5205.
- Newman, M. E. J. (2004b). Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):321–330.
- Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113.

- Pastor-Satorras, R., Vazquez, A., and Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, 87.
- Phelps, C. (2003). *Technological exploration: A longitudinal study of the role of recombinatory search and social capital in alliance networks*. PhD thesis, New York University, Graduate School of Business Administration.
- Powell, W., Koput, K., and Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, pages 116–145.
- Powell, W., White, D., Koput, K., and Owen-Smith, J. (2005a). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences¹. *American journal of sociology*, 110(4):1132–1205.
- Powell, W. W., White, D. R., Koput, K. W., and Owen-Smith, J. (2005b). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110(4):1132–1205.
- Ramasco, J. J., Dorogovtsev, S. N., and Pastor-Satorras, R. (2004). Self-organization of collaboration networks. *Physical Review E*, 70(3):036106.
- Reed, W. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74(1):15–19.
- Riccaboni, M. and Pammolli, F. (2002). On firm growth in networks. *Research Policy*, 31(8-9):1405–1416.
- Rodriguez-Girones, M. A. and Santamaria, L. (2006). A new algorithm to calculate the nestedness temperature of presence-absence matrices. *Journal of Biogeography*, 33(5):924–935.
- Rosenkopf, L. and Schilling, M. (2007). Comparing alliance network structure across industries: observations and explanations. *Strategic Entrepreneurship Journal*, 1(3-4):191–209.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.
- Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal*, 30:233–260.
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.
- Tiwana, A. (2008). Do bridging ties complement strong ties? an empirical examination of alliance ambidexterity. *Strategic Management Journal*, 29(3):251–272.
- Verspagen, B. and Duysters, G. (2004). The small worlds of strategic technology alliances. *Technovation*, 24(7):563–571.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393:440–442.

Westbrock, B. (2010). Natural concentration in industrial research collaboration. *The RAND Journal of Economics*, 41(2):351–371.